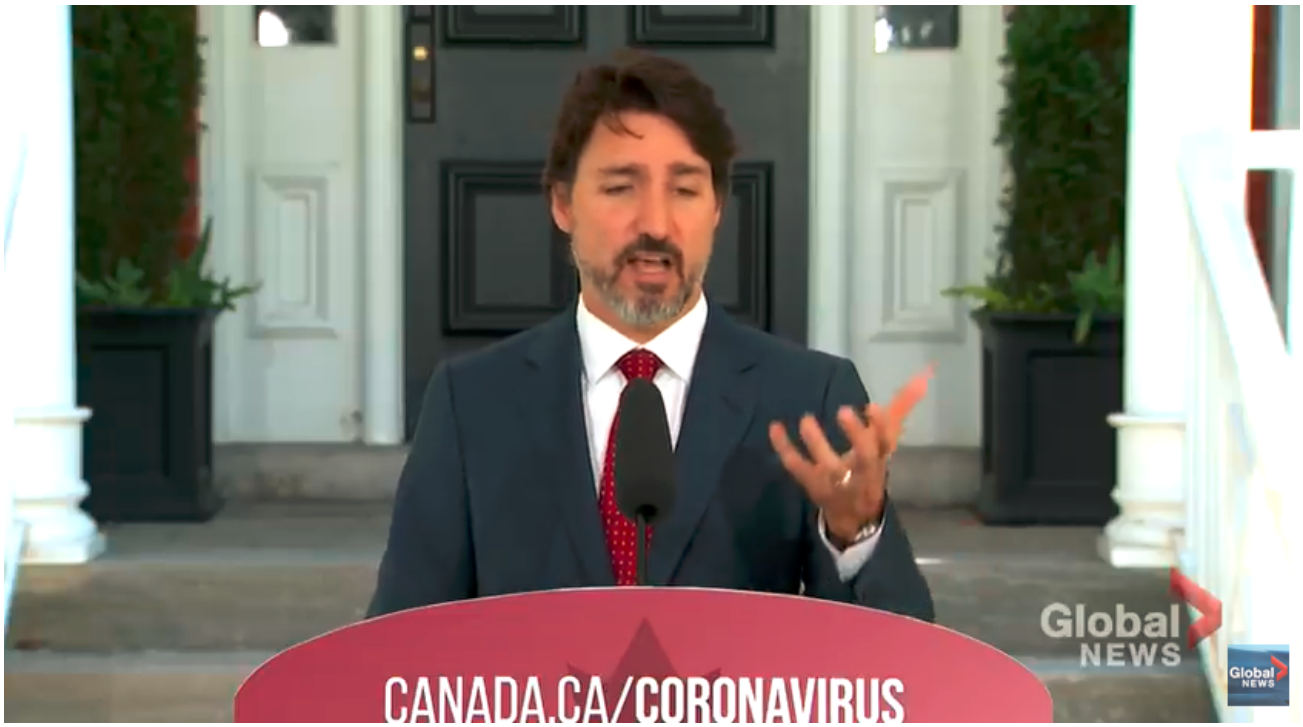
CANUCK LAW
Reporting That You're Not Getting Elsewhere (Do Your Own Research To Verify)

JUNE 28, 2020 BY RONNIE

De-Anonymizing The Anonymous Contact Tracing App

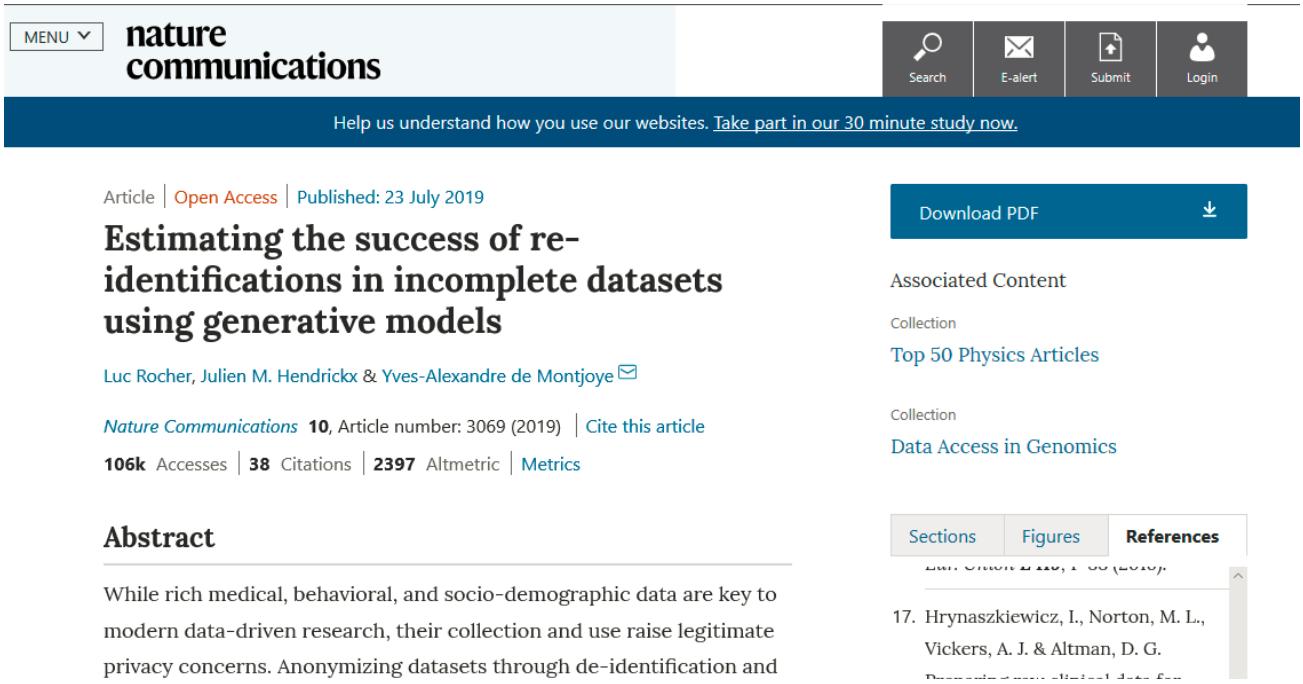


00:00 08:37

To start out with a disclaimer, I am hardly any sort of expert on cell phone technology. So this article is written from a more lay perspective. Nonetheless, the announcement of the contact tracing app in Canada opens up a lot of hard questions that need to be answered. Can the Government (or any government) be trusted with this claim, and is it even feasible?

This isn't meant to be an alarmist piece, but there are very real concerns and doubts about just how confidential all of this will remain. Consider the following.

Research Into Re-Identification, 2019


MENU nature communications

Search E-alert Submit Login

Help us understand how you use our websites. Take part in our 30 minute study now.

Article | Open Access | Published: 23 July 2019

Estimating the success of re-identifications in incomplete datasets using generative models

Luc Rocher, Julien M. Hendrickx & Yves-Alexandre de Montjoye

Nature Communications **10**, Article number: 3069 (2019) | Cite this article

106k Accesses | **38** Citations | **2397** Altmetric | Metrics

Abstract

While rich medical, behavioral, and socio-demographic data are key to modern data-driven research, their collection and use raise legitimate privacy concerns. Anonymizing datasets through de-identification and

Download PDF

Associated Content

Collection

Top 50 Physics Articles

Collection

Data Access in Genomics

Sections Figures References

17. Hrynaszkiewicz, I., Norton, M. L., Vickers, A. J. & Altman, D. G.

> While rich medical, behavioral, and socio-demographic data are key to modern data-driven research, their collection and use raise legitimate privacy concerns. Anonymizing

datasets through de-identification and sampling before sharing them has been the main tool used to address those concerns. We here propose a generative copula-based method that can accurately ***estimate the likelihood of a specific person to be correctly re-identified, even in a heavily incomplete dataset***. On 210 populations, our method obtains AUC scores for predicting individual uniqueness ranging from 0.84 to 0.97, with low false-discovery rate. Using our model, ***we find that 99.98% of Americans would be correctly re-identified in any dataset using 15 demographic attributes. Our results suggest that even heavily sampled anonymized datasets are unlikely to satisfy the modern standards for anonymization set forth by GDPR*** and seriously challenge the technical and legal adequacy of the de-identification release-and-forget model.

De-identification, the process of anonymizing datasets before sharing them, has been the main paradigm used in research and elsewhere to share data while preserving people's privacy. ***Data protection laws worldwide consider anonymous data as not personal data anymore allowing it to be freely used, shared, and sold***. Academic journals are, e.g., ***increasingly requiring authors to make anonymous data available to the research community***. While standards for anonymous data vary, modern data protection laws, such as the European General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), consider that ***each and every person in a dataset has to be protected for the dataset to be considered anonymous. This new higher standard for anonymization is further made clear by the introduction in GDPR of pseudonymous data: data that does not contain obvious identifiers but might be re-identifiable*** and is therefore within the scope of the law.

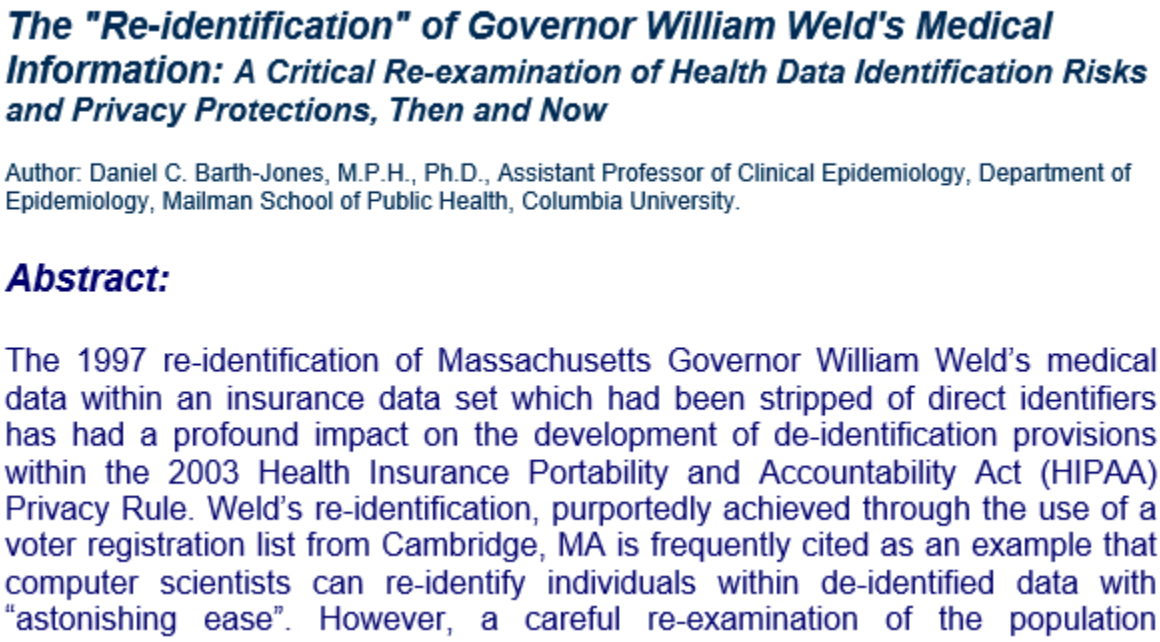
This was a research paper released in 2019, before the coronavirus planned-emic hit the world stage. While to long to into depth here, the researchers found and listed many examples of people being able to re-identify people using supposedly anonymized data sets. While original data had many modifiers removed, it was possible to reverse engineer it, and re-establish people's identities using multiple sets of incomplete data.

Two of the biggest issues in the research were health care data and internet browsing data. They were initially anonymized, but then computers were able to piece together to data and provide names. While not always correct, these techniques were overall very accurate in re-establishing identities.

Research data is widely shared for many purposes. Laws in the West allow for personal information to be shared as long as it is “anonymized” first. However, if that can be undone, then an end run around privacy laws can be accomplished.

Now, this type of bypass of privacy has been underway for a long time. People have to ask whether it will continue (or even escalate), in the face of this so-called pandemic.

Governor William Weld’s Medical Info



***The "Re-identification" of Governor William Weld's Medical Information:* A Critical Re-examination of Health Data Identification Risks and Privacy Protections, Then and Now**

Author: Daniel C. Barth-Jones, M.P.H., Ph.D., Assistant Professor of Clinical Epidemiology, Department of Epidemiology, Mailman School of Public Health, Columbia University.

Abstract:

The 1997 re-identification of Massachusetts Governor William Weld’s medical data within an insurance data set which had been stripped of direct identifiers has had a profound impact on the development of de-identification provisions within the 2003 Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. Weld’s re-identification, purportedly achieved through the use of a voter registration list from Cambridge, MA is frequently cited as an example that computer scientists can re-identify individuals within de-identified data with “astonishing ease”. However, a careful re-examination of the population

Re-Identification_of_Welds_Medical_Information

This is an old case, but a good one. Former Massachusetts Governor William Weld was able to have his medical history re-identified from anonymized medical information. How so? State voter rolls provided birth date and zip code information. Being a public figure, people knew quite a bit about him. Even with redacted records, it was possible to piece it together.

But one doesn’t have to be a politician. With the information available from various databases, a computer scientist can easily piece profiles together.

Keep in mind this was done in 1997, and led to HIPPA, new privacy regulations coming into

place. However, that was over 20 years ago, and computers have advanced a long way since. Moreover, internet usage has resulted in astronomical amounts of personal information being available online.

Now for some questions about this app.

5Will The App Really Be Anonymous?

The first thing that people should be asking is whether claims that this app will be anonymous at all. A healthy distrust of the your government is helpful in all cases. Everything they say and promise should be met with some degree of skepticism.

Bear in mind, this is the same government that thought nothing of having Statistics Canada do data mining of over 500,000 Canadians. They then threw StatsCan under the bus when there was public backlash. It was just 2 years ago, and addressed in those articles.

Beyond distrust of the government, a follow-up must be asked. Even if this were anonymous, as advertised, can it be de-anonymized at a later point? Can the app makers use some decryption to identify users? What about other third parties?

How easy will it be to use AI or to combine partial data sets to re-identify people? What happens when the profiles are “Frankenstein-ed” together? Who gets the data? How will it be used, and will we even know?

What Qualifies As Contact?

Is passing someone on the street or in the grocery store sufficient to count as “coming in contact” with someone? is a few seconds enough? A minute? 5 minutes? Sure there is more information coming out, but having some standard would be nice. Knowing what the standard is would also help.

Positive Test Linked To Phone Number?

There are plenty of issues with the coronavirus testing itself. However, that is a piece for another day. This is about the privacy aspects.

Suppose you test positive for this virus. What happens then? Do you change the settings on your phone, or does the medical staff then insert your phone number or “random number” into a database of people who have tested positive? Is that result then connected to anything and anyplace you go, or that your phone is reported to have a connection to?

Lies About Phone Not Geo-Tagging?

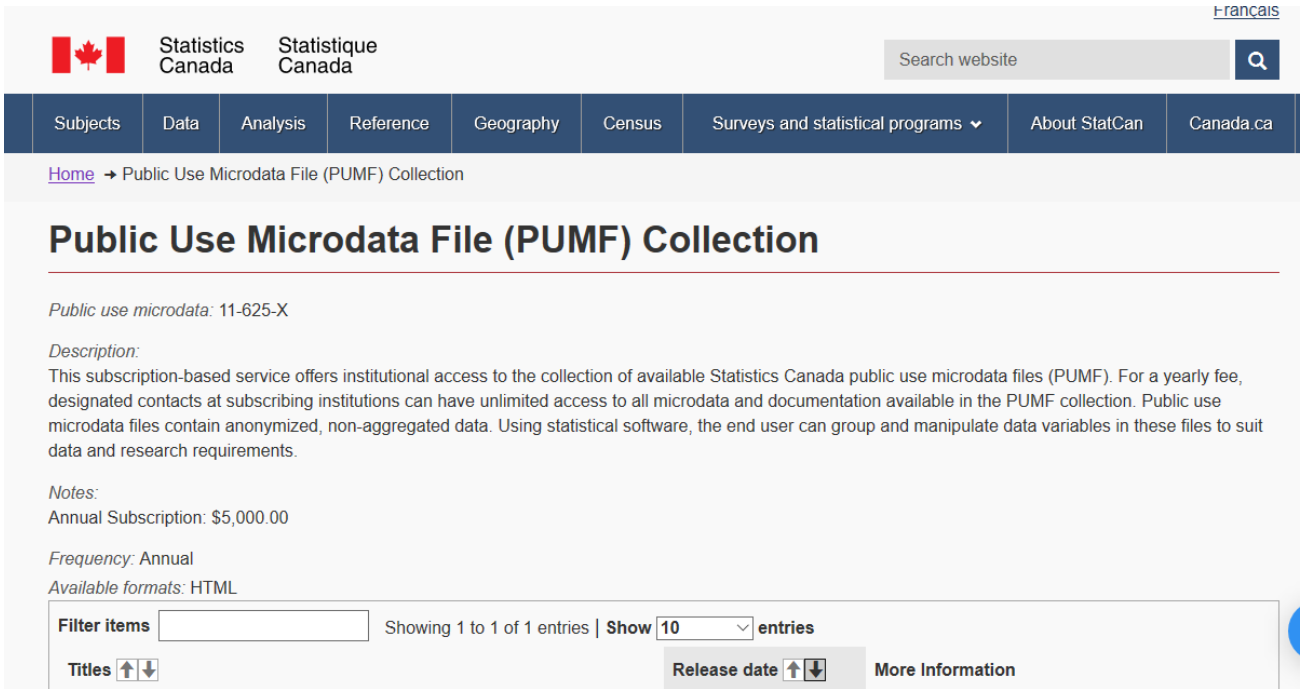
There are claims that there will be no geo-tagging, or storing of locations. How exactly does that work though? How can a phone app determine that a user has been close to someone who has tested positive? It’s difficult to believe that phones would just start collecting the random assigned numbers of everyone it has been close to (though possible I guess), but not record any sort of geographical data?

Any sort of mainstream technology that has GPS tracking can find places, people or things, but does so with reference to spots on a map. How could this contact tracing app determine when phones are close to each other, but not have any geographical reference?

It seems possible that this government app could use geographical references, but then not store the data. However, considering outfits like Google are well ahead in tracking movements, it seems strange to develop this app to not record location data.

StatsCan Provides Microdata For Free

Français
Statistics Canada
Statistique Canada
Search website
Subjects
Data
Analysis
Reference
Geography
Census
Surveys and statistical programs
About StatCan
Canada.ca
Home → Access to microdata
Access to microdata
Statistics Canada recognizes that researchers require access to microdata at the individual business, household or person level for research purposes. To preserve the privacy and confidentiality of respondents, and to encourage the use of microdata, Statistics Canada offers a wide range of options through a series of online channels, facilities and programs.
CONTINUUM of MICRODATA ACCESS
Nonrestricted Access
Restricted Access
DATA LIBERATION INITIATIVE (DLI)
ACCESS TO PUBLIC USE MICRODATA FILES (PUMFS) COLLECTION
REAL TIME REMOTE ACCESS (RTRA)
RESEARCH DATA CENTRES (RDC)
CANADIAN CENTRE FOR DATA DEVELOPMENT AND ECONOMIC RESEARCH (CDER)
Canada

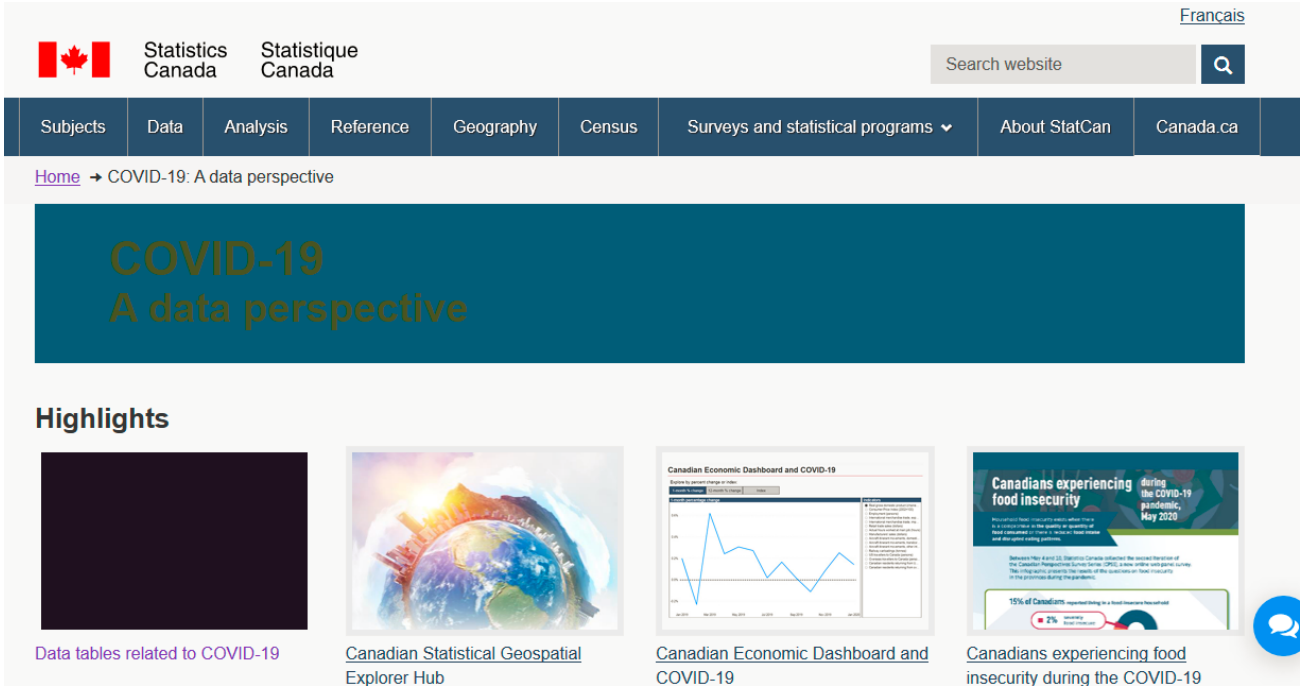

Français
Statistics Canada
Statistique Canada
Search website
Subjects
Data
Analysis
Reference
Geography
Census
Surveys and statistical programs
About StatCan
Canada.ca
Home → Public Use Microdata File (PUMF) Collection
Public Use Microdata File (PUMF) Collection
Public use microdata: 11-625-X
Description:
This subscription-based service offers institutional access to the collection of available Statistics Canada public use microdata files (PUMF). For a yearly fee, designated contacts at subscribing institutions can have unlimited access to all microdata and documentation available in the PUMF collection. Public use microdata files contain anonymized, non-aggregated data. Using statistical software, the end user can group and manipulate data variables in these files to suit data and research requirements.
Notes:
Annual Subscription: $5,000.00
Frequency: Annual
Available formats: HTML
Filter items
Showing 1 to 1 of 1 entries | Show 10 entries
Titles
Release date
More Information

Unrestricted access to microdata

Statistics Canada offers Public Use Microdata Files (PUMFs) to institutions and individuals. They are non-aggregated data which are ***carefully modified and then reviewed to ensure that no individual or business is directly or indirectly identified***. These can be accessed directly through the Data Liberation Initiative (DLI) or the PUMF Collection for a subscription fee. ***Individual files can also be requested at no cost***.

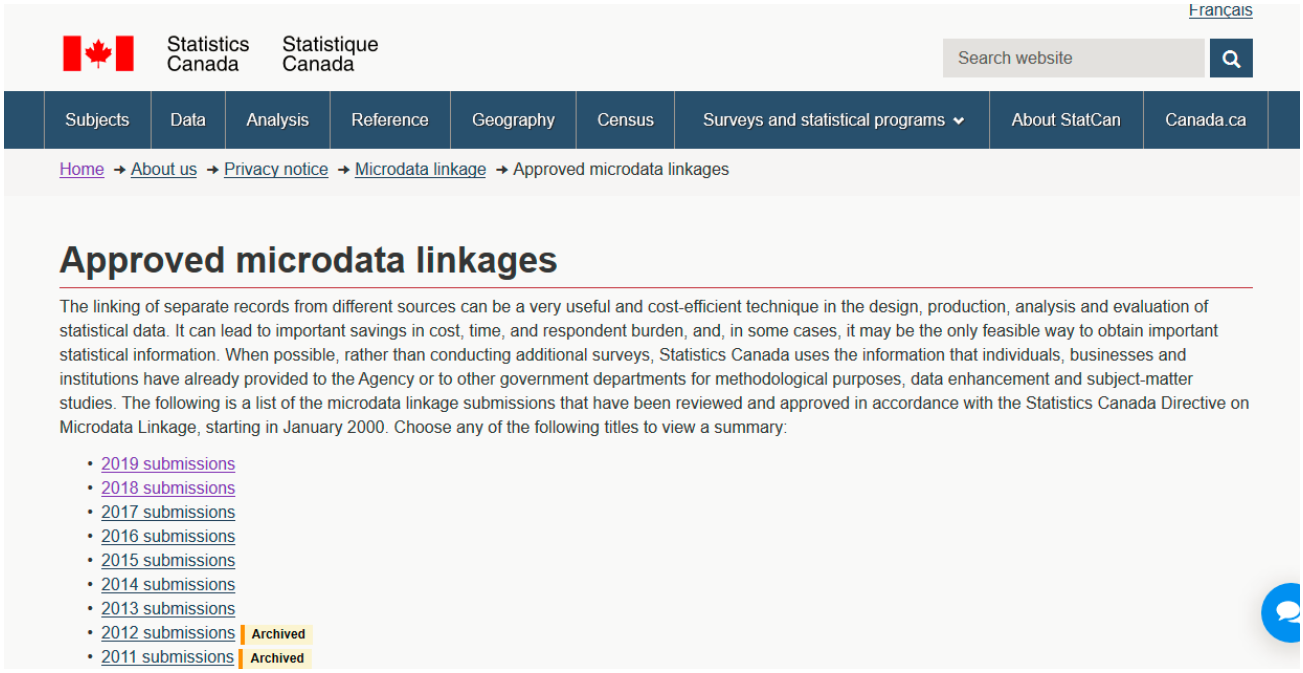
For reference, a files can be ordered for free. A purchase of $5,000 per year, which gives

unlimited access to all of the microdata used by StatCan in its various research and publications. The data is supposed to be anonymized, but one has to ask how easy it would be to piece together individual or businesses, based on this information, plus other available sources.



StatsCan already has plenty of CV-19 research released and available for the public. It isn't too much of a stretch to think that searching for where people cluster, or amount of time spent in an area is researched.

StatsCan's "Approved Microdata Linkages"



Français

Statistics Canada Statistique Canada

Search website

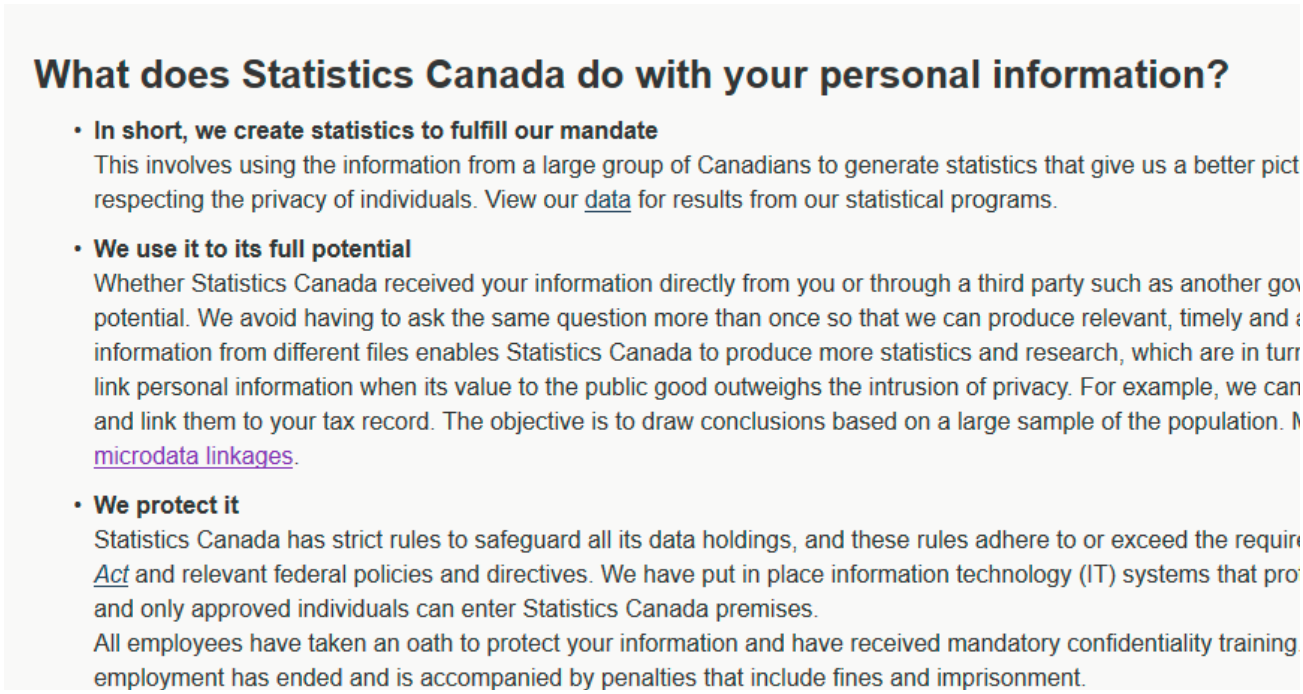
Subjects | Data | Analysis | Reference | Geography | Census | Surveys and statistical programs | About StatCan | Canada.ca

Home → About us → Privacy notice → Microdata linkage → Approved microdata linkages

Approved microdata linkages

The linking of separate records from different sources can be a very useful and cost-efficient technique in the design, production, analysis and evaluation of statistical data. It can lead to important savings in cost, time, and respondent burden, and, in some cases, it may be the only feasible way to obtain important statistical information. When possible, rather than conducting additional surveys, Statistics Canada uses the information that individuals, businesses and institutions have already provided to the Agency or to other government departments for methodological purposes, data enhancement and subject-matter studies. The following is a list of the microdata linkage submissions that have been reviewed and approved in accordance with the Statistics Canada Directive on Microdata Linkage, starting in January 2000. Choose any of the following titles to view a summary:

- 2019 submissions
- 2018 submissions
- 2017 submissions
- 2016 submissions
- 2015 submissions
- 2014 submissions
- 2013 submissions
- 2012 submissions Archived
- 2011 submissions Archived



What does Statistics Canada do with your personal information?

- **In short, we create statistics to fulfill our mandate**
  This involves using the information from a large group of Canadians to generate statistics that give us a better pict
  respecting the privacy of individuals. View our data for results from our statistical programs.
- **We use it to its full potential**
  Whether Statistics Canada received your information directly from you or through a third party such as another gov
  potential. We avoid having to ask the same question more than once so that we can produce relevant, timely and a
  information from different files enables Statistics Canada to produce more statistics and research, which are in turn
  link personal information when its value to the public good outweighs the intrusion of privacy. For example, we can
  and link them to your tax record. The objective is to draw conclusions based on a large sample of the population. M
  microdata linkages.
- **We protect it**
  Statistics Canada has strict rules to safeguard all its data holdings, and these rules adhere to or exceed the require
  Act and relevant federal policies and directives. We have put in place information technology (IT) systems that pro
  and only approved individuals can enter Statistics Canada premises.
  All employees have taken an oath to protect your information and have received mandatory confidentiality training
  employment has ended and is accompanied by penalties that include fines and imprisonment.

What does Statistics Canada do with your personal information?

.

We use it to its full potential

Whether Statistics Canada received your information directly from you or through a third party such as another government entity, ***we use it to its full potential***. We avoid having to ask the same question more than once so that we can produce relevant, timely and accurate statistics. ***Linking Canadians' information from different files enables Statistics Canada to produce more statistics and research, which are in turn used***

by decision makers. We will only link personal information when its value to the public good outweighs the intrusion of privacy. For example, we can take the answers you gave on a survey and link them to your tax record. The objective is to draw conclusions based on a large sample of the population. More information on all Approved microdata linkages.

StatsCan openly admits that it will combine data from various sources and combine it. So this “anonymizing” is only done AFTER various things are combined, if it even done at all.

Approved microdata linkages

.

The ***linking of separate records from different sources can be a very useful and cost-efficient technique in the design, production, analysis and evaluation of statistical data***. It can lead to important savings in cost, time, and respondent burden, and, in some cases, ***it may be the only feasible way to obtain important statistical information. When possible, rather than conducting additional surveys, Statistics Canada uses the information that individuals, businesses and institutions have already provided to the Agency or to other government departments for methodological purposes, data enhancement and subject-matter studies.*** The following is a list of the microdata linkage submissions that have been reviewed and approved in accordance with the ***Statistics Canada Directive on Microdata Linkage***, starting in January 2000. Choose any of the following titles to view a summary:

To be clear, Statistics Canada already has the system of combining various datasets (including information provided by other government agencies, schools, businesses and institutions. In fact, it has gone this for a good 20 years now. Presumably the anonymising is done ***AFTER*** this is compiled.

Looking at the approved microdata linking from 2019 (the most recent year), we get:

- Evaluating the Information Content in the Business Outlook Survey (002-2019)
- Evaluating the Information Content in the Business Outlook Survey (002-2019)
- The impact of Intellectual Property on the Canadian Economy (003-2019)
- The impact of Intellectual Property on the Canadian Economy (003-2019)
- LASS 2016 to Census 2016, Census 2011 and NHS 2011 Linkage (004-2019)

- LASS 2016 to Census 2016, Census 2011 and NHS 2011 Linkage (004-2019)
- Linkage of the National Dose Registry to cancer and mortality outcomes, an update (005-2019)
- Linkage of the National Dose Registry to cancer and mortality outcomes, an update (005-2019)
- Municipal Wastewater Systems in Canada (MWSC): Environment and Climate Change Canada (ECCC) Effluent Regulatory Reporting Information System (ERRIS) linkage to Census Data (006-2019)
- Municipal Wastewater Systems in Canada (MWSC): Environment and Climate Change Canada (ECCC) Effluent Regulatory Reporting Information System (ERRIS) linkage to Census Data (006-2019)
- Adding Gender to the Corporations Returns Act (CRA) database (007-2019)
- Adding Gender to the Corporations Returns Act (CRA) database (007-2019)
- Between and within-firm earnings inequality in Canada (008-2019)
- Between and within-firm earnings inequality in Canada (008-2019)
- Indian Register linked to tax data, (Longitudinal Indian Register Database (LIRD)) (009-2019)
- Indian Register linked to tax data, (Longitudinal Indian Register Database (LIRD)) (009-2019)
- 2016 Census of Population linkage to income tax files and benefits records to monitor tax filing behaviour and take-up rate of various benefit programs (011-2019)
- 2016 Census of Population linkage to income tax files and benefits records to monitor tax filing behaviour and take-up rate of various benefit programs (011-2019)
- Linkage of the 2002 Canadian Community Health Survey – Mental Health and Well-being – Canadian Forces (CCHS-CF) to the 2018 Canadian Armed Forces Members and Veterans Mental Health Follow-up Survey (CAFVMHS) (021-2019)
- Linkage of the 2002 Canadian Community Health Survey – Mental Health and Well-being – Canadian Forces (CCHS-CF) to the 2018 Canadian Armed Forces Members and Veterans Mental Health Follow-up Survey (CAFVMHS) (021-2019)
- Socioeconomic and Ethnocultural Disparities in Perinatal Health in Canada: Current Pattern and Changes Over Time (023-2019)
- Socioeconomic and Ethnocultural Disparities in Perinatal Health in Canada: Current Pattern and Changes Over Time (023-2019)
- Linkage of the Canadian Housing Survey to historical income information, information on social and affordable housing, measures on proximity to services and measures on income

dispersion in communities (024-2019)

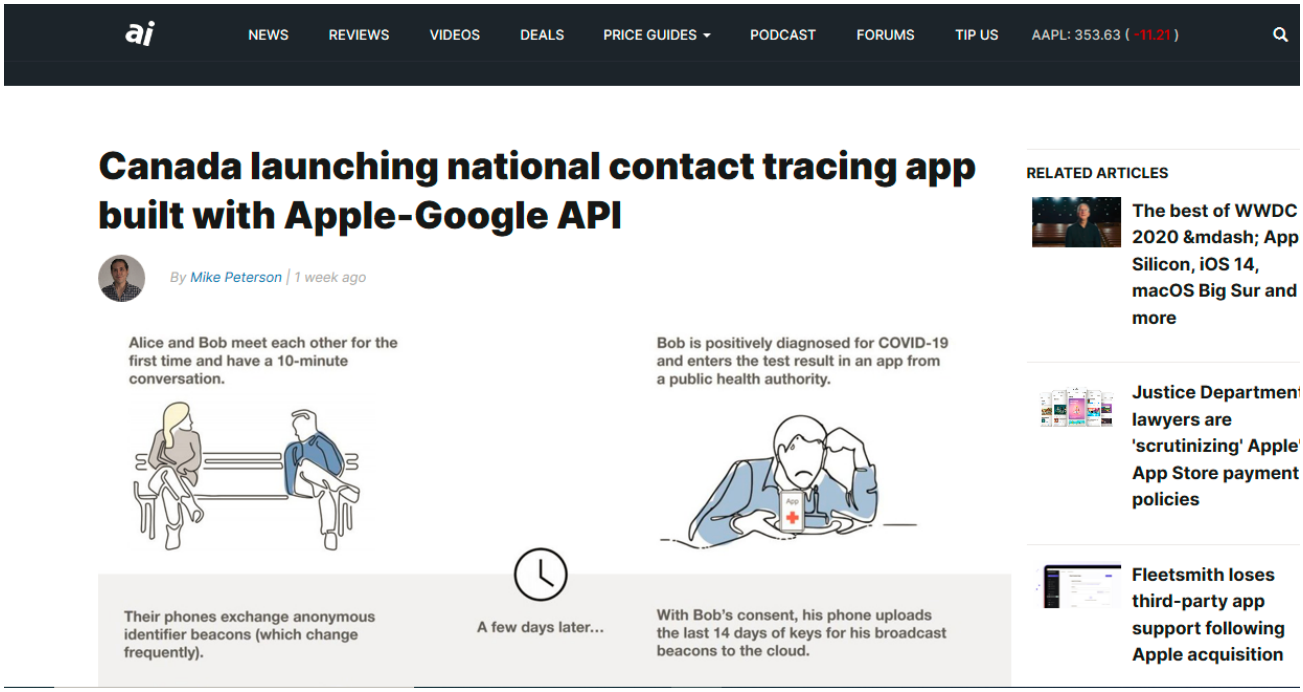
- Linkage of the Canadian Housing Survey to historical income information, information on social and affordable housing, measures on proximity to services and measures on income dispersion in communities (024-2019)
- Linkage of Labour Force Survey with Longitudinal Workers File (025-2019)
- Linkage of Labour Force Survey with Longitudinal Workers File (025-2019)
- The Economic and Environmental Impacts of Voluntary Energy Conservation Programs: Evidence from the Canadian Industry Program for Energy Conservation (026-2019)
- The Economic and Environmental Impacts of Voluntary Energy Conservation Programs: Evidence from the Canadian Industry Program for Energy Conservation (026-2019)

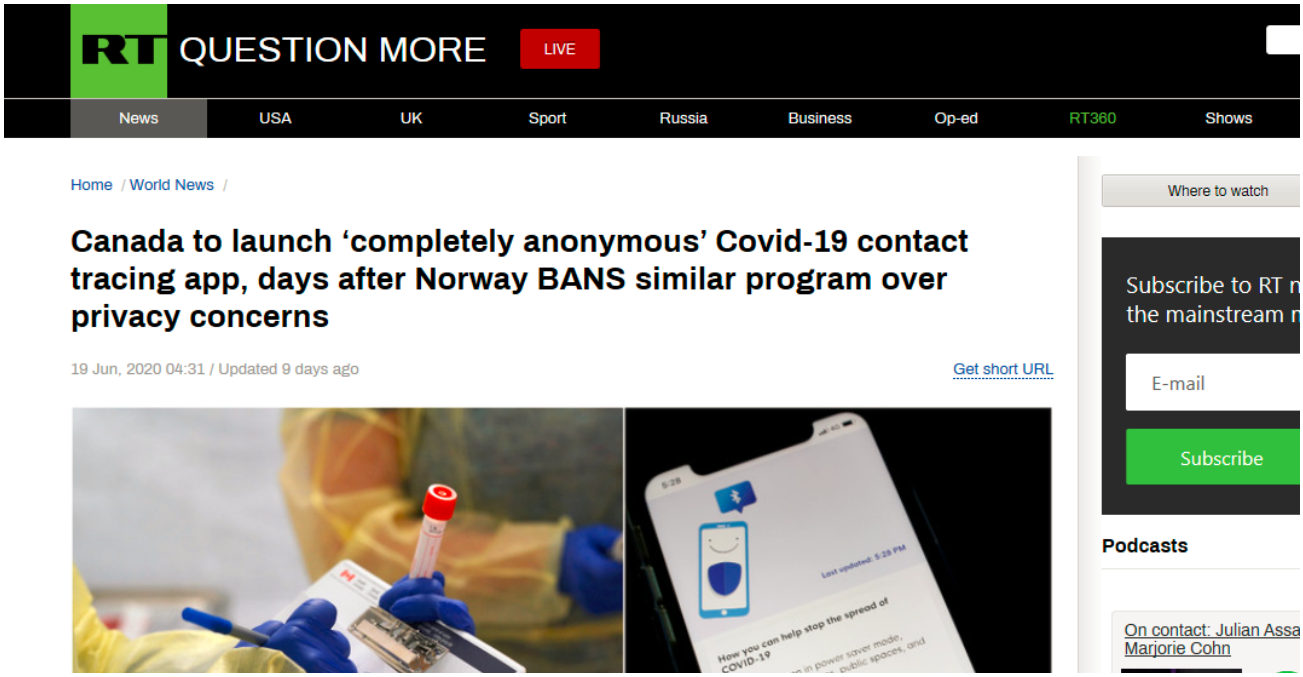
Since Statistics Canada already incorporates health information and combines various sets of data to make “more complete profiles”, it is clearly possible to add CV tests — both positive and negative as well. While calling for it publicly is political poison, who’s to say it won’t be quietly slipped in at some point?

Remember as well, these profiles are combined, and only then anonymized. However, the more information in the profile, the easier it would be for researchers to reverse engineer the anonymizing techniques to restore identities. In fact, it’s quite possible that the algorithm and techniques will be readily available.

Remember, StatsCan allows people to order individual files for free. It you want a full 1-year subscription, it costs a mere $5,000. If you are interested in real data mining, it’s pocket change.

Shopify & Blackberry Develop App


ai
NEWS
REVIEWS
VIDEOS
DEALS
PRICE GUIDES
PODCAST
FORUMS
TIP US
AAPL: 353.63 (-11.21)
Canada launching national contact tracing app built with Apple-Google API
By Mike Peterson | 1 week ago
Alice and Bob meet each other for the first time and have a 10-minute conversation.
Bob is positively diagnosed for COVID-19 and enters the test result in an app from a public health authority.
Their phones exchange anonymous identifier beacons (which change frequently).
A few days later...
With Bob's consent, his phone uploads the last 14 days of keys for his broadcast beacons to the cloud.
RELATED ARTICLES
The best of WWDC 2020 — App Silicon, iOS 14, macOS Big Sur and more
Justice Department lawyers are 'scrutinizing' Apple's App Store payment policies
Fleetsmith loses third-party app support following Apple acquisition



Canada will launch a nationwide contact tracing app using the Apple-Google Exposure Notification framework, Prime Minister Justin Trudeau said Thursday.

The ***Apple-Google Exposure Notification API exited beta in May. It allows public health authorities to build deeply integrated, cross-platform contact tracing apps*** to track and curb the spread of coronavirus.

> The Canadian app was ***developed by Shopify, BlackBerry and the government of Ontario. As is required by Apple and Google, the app will be completely voluntary***, will only store data in a decentralized manner and will be led by the Canadian Digital Service Initiative, iPhoneInCanada reported.

Blackberry and Shopify developed the app for use in Canada. Companies like Google are well known for obtaining huge amounts of data on their users so this is a huge red flag. How do we know there isn't some sort of back door built into the platform?

By contrast, a few countries, like Norway, have banned such an app, out of privacy concerns.

Government Already Compiles The Info

Statistics Canada
Statistique Canada
Search website
Subjects
Data
Analysis
Reference
Geography
Census
Surveys and statistical programs
About StatCan
Canada.ca
Home → About us → Privacy impact assessments → Generic Privacy Impact Assessment for Statistics Canada's Statistical Programs → Acquisition of credit information
Acquisition of credit information
Supplement to Statistics Canada's Generic Privacy Impact Assessment related to the acquisition of credit information
Date: March 2018
Program manager: Director, Administrative Data Division
Reference to Personal Information Bank:
In accordance with the Privacy Act, Statistics Canada is developing a new institutional personal information bank (PIB) to describe the personal credit and financial information Statistics Canada has obtained from public and private sector sources for the purposes of the Statistics Act, including but not limited to the following programs: Financial and Wealth Accounts, Distributed Household Economic Accounts, Canadian Housing, and Income Statistics Programs. Amongst

As seen in earlier sections, StatsCan already combines sources to build "more complete" profiles of the people it wants to survey. Even your credit isn't safe if StatsCan wants it. As for the finished project, the information can be bought, and individual files requested for free. How difficult would it be to take the raw data provided, and cross reference across other social media or other databases? How long until the original names are restored to the profiles?

With all this data compilation, it won't be difficult to link a positive test to a real name, an address, or a date of birth. The suggestion that all of this will remain completely anonymous flies in the face of what the government and StatsCan do.

It also isn't much of a stretch to see the "anonymized" results sold or given to third parties to conduct their own research. Stay away from the app would be some good advice. It would be nice to just take at face value the claims that there are no privacy issues. However, that's very naïve.

Again, this is not meant to send people into a panic, but much more has to be known and discussed to make such an app a real solution, if it is at all.

BIG TECH, CANADA, PRIVACY, VACCINES, WARS

One Reply to "De-Anonymizing The Anonymous Contact Tracing App"

cecil1

JUNE 28, 2020 AT 1:31 PM

Outstanding work here.

And very bloody disturbing.

Rule #1: Never trust government.

Rule #2: Never give government the power so you have to 'trust' them.